

# The Advantage of Truncated Permutations

Shoni Gilboa\*

Shay Gueron<sup>†‡</sup>

October 11, 2016

## Abstract

Let  $m < n$  be non-negative integers. An oracle chooses a permutation  $\pi$  of  $\{0, 1\}^n$  uniformly at random. When queried with an  $n$ -bit string  $w$ , it truncates the last  $m$  bits of  $\pi(w)$ , and returns the remaining first  $n - m$  bits. Such truncated random permutations were suggested by Hall et al., in 1998, as a construction of a Pseudo Random Function. They conjectured that the distinguishing advantage of this PRF, given a budget of  $q$  queries,  $\mathbf{Adv}_{n,m}(q)$ , is small if  $q = o(2^{(m+n)/2})$ . They established a general upper bound on  $\mathbf{Adv}_{n,m}(q)$ , which confirms the conjecture only for  $m < n/7$ . The conjecture was essentially confirmed by Bellare and Impagliazzo in 1999. Nevertheless, the problem of estimating  $\mathbf{Adv}_{n,m}(q)$  remained open.

Combining the trivial bound 1, the birthday bound, and a result that Stam had published much earlier in 1978, in a different context, leads to the following upper bound:

$$\mathbf{Adv}_{n,m}(q) = O\left(\min\left\{\frac{q^2}{2^n}, \frac{q}{2^{\frac{n+m}{2}}}, 1\right\}\right)$$

This paper settles the open problem by showing that this bound is tight.

**Keywords:** Pseudo random permutations, pseudo random functions, advantage.

## 1 Introduction

Distinguishing a random permutation from a random function is a combinatorial problem which has a fundamental role in cryptology. Analyses of cryptographic primitives, such as block ciphers, hash and MAC schemes, often start with an idealization of the primitive as a uniform random permutation, in order to derive information theoretic security bounds. In this paper, we discuss a generalization of this problem.

We first recall several concepts which are frequently used in cryptography. Let  $\ell, n$  be positive integers and let  $\mathcal{F}_{n,\ell}$  be the set of functions from  $\{0, 1\}^n$  to  $\{0, 1\}^\ell$ . A Pseudo Random Function

---

\*Department of Mathematics and Computer Science, The Open University of Israel, Raanana 4353701, Israel. [tipshoni@gmail.com](mailto:tipshoni@gmail.com)

<sup>†</sup>Department of Mathematics, University of Haifa, Haifa 3498838, Israel. [shay@math.haifa.ac.il](mailto:shay@math.haifa.ac.il)

<sup>‡</sup>Intel Corporation, Israel Development Center, Haifa, 31015, Israel.

(PRF)  $\Phi : \{0,1\}^n \rightarrow \{0,1\}^\ell$  is a selection of a function from  $\mathcal{F}_{n,\ell}$ , according to some probability distribution. The quality of a PRF  $\Phi$  is determined by the ability of an “adversary” to distinguish an instance of  $\Phi$  from a function chosen uniformly at random from  $\mathcal{F}_{n,\ell}$ , in the following setting. It is assumed that the adversary has only query access to a function  $\varphi : \{0,1\}^n \rightarrow \{0,1\}^\ell$ , which is either selected uniformly at random from  $\mathcal{F}_{n,\ell}$ , or is an instance of the PRF  $\Phi$ . The adversary may use any algorithm  $\mathcal{A}$  that first selects (possibly adaptively) a sequence of queries to the functions, i.e. strings in  $\{0,1\}^n$ , and then, based on the answers to the queries, outputs a bit. We may interpret this bit as the guess of  $\mathcal{A}$ . For  $b \in \{0,1\}$ , let  $P_\Phi^{\mathcal{A}}(b)$  be the probability that the output is  $b$  when  $\varphi$  is our PRF, and let  $P_U^{\mathcal{A}}(b)$  be the probability that the output is  $b$  when  $\varphi$  is selected from  $\mathcal{F}_{n,\ell}$  uniformly at random. The advantage of  $\mathcal{A}$  against the PRF  $\Phi$  is defined as  $|P_\Phi^{\mathcal{A}}(1) - P_U^{\mathcal{A}}(1)|$  (which also equals  $|P_\Phi^{\mathcal{A}}(0) - P_U^{\mathcal{A}}(0)|$ ). The advantage,  $\mathbf{Adv}_\Phi$ , of the adversary against the PRF  $\Phi$ , is the maximal advantage of  $\mathcal{A}$  against  $\Phi$  over all the algorithms he may use, as a function of the number of queries. Hereafter, we consider adversaries with no computational limitations, in which case the advantage has an explicit expression which is presented in Section 2.

**The permutation based PRF.** A classical example of a PRF from  $\{0,1\}^n$  to  $\{0,1\}^n$  is a permutation of  $\{0,1\}^n$ , chosen uniformly at random. In this case, the advantage of the PRF can be computed by means of the simple “collision test” and the Birthday Problem, and is

$$\mathbf{Adv}(q) = 1 - \left(1 - \frac{1}{2^n}\right) \left(1 - \frac{2}{2^n}\right) \dots \left(1 - \frac{q-1}{2^n}\right).$$

In particular,  $\mathbf{Adv}(q) = 1$  for  $q > 2^n$ . For  $q \leq 2^n$ , the following approximations hold. Since for every  $1 \leq k \leq q-1$  we have

$$1 - \frac{q}{2^n} \leq \left(1 - \frac{k}{2^n}\right) \left(1 - \frac{q-k}{2^n}\right) \leq \left(1 - \frac{q}{2^{n+1}}\right)^2,$$

it follows that

$$1 - e^{-\frac{q(q-1)}{2^{n+1}}} \leq 1 - \left(1 - \frac{q}{2^{n+1}}\right)^{q-1} \leq \mathbf{Adv}(q) \leq 1 - \left(1 - \frac{q}{2^n}\right)^{\frac{q-1}{2}} \leq \min \left\{ \frac{q(q-1)}{2^{n+1}}, 1 \right\}.$$

Therefore,

$$\mathbf{Adv}(q) = \Theta \left( \min \left\{ q^2/2^n, 1 \right\} \right). \quad (1)$$

This result implies that the number of queries required to distinguish a random permutation from a random function, with success probability significantly larger than, say,  $1/2$ , is  $\Theta(2^{n/2})$ .

A generalization of the above PRF is the following.

**Definition 1** (Truncated Permutation PRF). *Let  $\mathbf{TRUNC}_{n,m} : \{0,1\}^n \rightarrow \{0,1\}^{n-m}$  be defined by the mapping  $(x_1, x_2, \dots, x_n) \mapsto (x_1, x_2, \dots, x_{n-m})$ . The “Truncated permutation” PRF is the PRF defined by the composition  $\mathbf{TRUNC}_{n,m} \circ \pi$ , where  $\pi$  is a permutation of  $\{0,1\}^n$ , chosen uniformly at random.*

**Notation.** The advantage of an (computationally unbounded) adversary against the Truncated Permutation PRF is denoted by  $\mathbf{Adv}_{n,m}$ .

The following problem arises naturally.

**Problem 1.** For every  $0 \leq m < n$  and  $q$ , find (the order of magnitude of)  $\mathbf{Adv}_{n,m}(q)$ .

A different, related, problem is the following.

**Problem 2.** For every  $0 \leq m < n$ , how many queries does the adversary need in order to gain non-negligible advantage against the Truncated Permutation PRF? Specifically, what is (the order of magnitude) of  $q_{1/2}(n, m) = \min\{q \mid \mathbf{Adv}_{n,m}(q) \geq 1/2\}$ ?

We proceed to describe the history of these problems.

**The Birthday bound (folklore).** We start by remarking that the classical "Birthday" bounds

$$\mathbf{Adv}_{n,m}(q) \leq \min \left\{ \frac{q(q-1)}{2^{n+1}}, 1 \right\}, \quad (2)$$

and  $q_{1/2}(n, m) = \Omega(2^{n/2})$  are obviously valid. In fact, any algorithm that the adversary can use with the truncated replies of  $(n - m)$  bits from  $\pi(w)$  ( $w \in \{0, 1\}^n$ ) can also be used by the adversary who sees the full  $\pi(w)$  (he can ignore  $m$  bits and apply the same algorithm). Of course, we expect better bounds that would reflect the fact that the adversary receives less information when  $f(w)$  is truncated.

**Hall et al. (1998).** Problems 1 and 2 were studied by Hall et al. [4] in 1998, where the truncated (random) permutation were proposed as a PRF construction. The authors declared the lower bound

$$\mathbf{Adv}_{n,m}(q) = \Omega(q^2/2^{n+m}), \quad (3)$$

for every  $0 \leq m < n$  and  $q \leq 2^{(n+m)/2}$ . This bound implies that  $q_{1/2}(n, m) = O(2^{(n+m)/2})$  for every  $0 \leq m < n$ . They also proved the following upper bound:

$$\mathbf{Adv}_{n,m}(q) \leq 5 \left( \frac{q}{2^{\frac{n+m}{2}}} \right)^{\frac{2}{3}} + \frac{1}{2} \left( \frac{q}{2^{\frac{n+m}{2}}} \right)^3 \frac{1}{2^{\frac{n-7m}{2}}} \quad (4)$$

For  $m \leq n/7$  this implies that  $q_{1/2}(n, m) = \Omega(2^{(m+n)/2})$ . However, for larger values of  $m$ , the bound on  $q_{1/2}(n, m)$  that is offered by (4) deteriorates, and becomes (already for  $m > n/4$ ) worse than the trivial "Birthday" bound  $q_{1/2}(n, m) = \Omega(2^{n/2})$ . Hall et al. [4] conjectured that an adversary needs  $\Omega(2^{(n+m)/2})$  queries in order to get a non-negligible advantage, in the general case.

**Bellare and Impagliazzo (1999).** Bellare and Impagliazzo derived the following result in 1999 [1, Theorem 4.2].

$$\mathbf{Adv}_{n,m}(q) = O(n) \frac{q}{2^{\frac{n+m}{2}}} \quad (5)$$

whenever  $2^{n-m} < q < 2^{\frac{n+m}{2}}$ . This implies that  $q_{\frac{1}{2}} = \Omega(\frac{1}{n} 2^{\frac{m+n}{2}})$  for  $m > \frac{1}{3}n + \frac{2}{3}\log_2 n + \Omega(1)$ .

**Gilboa and Gueron (2015).** The method used to show (4) can be pushed to prove the conjecture made in [4], thus settling Problem 2, for almost every  $m$ . In particular, [2] showed that

$$\mathbf{Adv}_{n,m}(q) \leq \begin{cases} 2\sqrt[3]{2} \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{\frac{2}{3}} + \frac{2\sqrt{2}}{\sqrt{3}} \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{\frac{3}{2}} + \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 & m \leq \frac{n}{3} \\ 3 \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^{\frac{2}{3}} + 2 \left(\frac{q}{2^{\frac{n+m}{2}}}\right) + 5 \left(\frac{q}{2^{\frac{n+m}{2}}}\right)^2 + \frac{1}{2} \left(\frac{2q}{2^{\frac{n+m}{2}}}\right)^{\frac{n}{n-m}} & \frac{n}{3} < m \leq n - \log_2(16n) \end{cases} \quad (6)$$

which implies that  $q_{1/2}(n, m) = \Omega(2^{\frac{m+n}{2}})$  for every  $0 \leq m \leq n - \log_2(16n)$ .

**Stam (1978).** Surprisingly, it turns out that Problem 2 was solved 20 years before Hall et al. [4], in a different context. The bound

$$\mathbf{Adv}_{n,m}(q) \leq \frac{1}{2} \sqrt{\frac{(2^{n-m} - 1)q(q-1)}{(2^n - 1)(2^n - (q-1))}} \leq \frac{1}{2\sqrt{1 - \frac{q-1}{2^n}}} \cdot \frac{q}{2^{\frac{n+m}{2}}}, \quad (7)$$

which is valid for every  $0 \leq m < n$  and  $q \leq 2^n$ , follows directly from a result of Stam [5, Theorem 2.3]. This implies that  $q_{1/2}(n, m) = \Omega(2^{(m+n)/2})$  for every  $0 \leq m < n$ , confirming the conjecture of [4] in all generality (20 years before the conjecture was raised). We point out that the bound in [5] can be simplified to the more amenable form

$$\mathbf{Adv}_{n,m}(q) \leq \frac{q}{2^{\frac{m+n}{2}}}, \quad q \leq \frac{3}{4}2^n \quad (8)$$

This settles Problem 2, but note that Problem 1 still remains quite open.

**The best known bounds for Problem 1.** Note that the bound (7) is tighter than the bounds (4), (5) and (6). Therefore, summarizing the above results, the best known upper bound for the advantage in Problem 1, is the one obtained by combining (2) and (7), namely

$$\mathbf{Adv}_{n,m}(q) \leq \min \left\{ \frac{q(q-1)}{2^{n+1}}, \frac{1}{2} \sqrt{\frac{(2^{n-m} - 1)q(q-1)}{(2^n - 1)(2^n - (q-1))}}, 1 \right\} = \Theta \left( \min \left\{ \frac{q^2}{2^n}, \frac{q}{2^{\frac{n+m}{2}}}, 1 \right\} \right), \quad (9)$$

whereas the only general lower bound that we are aware of is the bound (3), declared in [4]. By (1), we know that the bound (9) is tight if  $m = 0$ , and it was shown in [3] that it is tight also in the case  $m = n - 1$ .

**Our contribution.** In this paper we answer Problem 1 by showing that (9) is tight for every  $q > 1$ , as formulated in the following theorem.

**Theorem 1.** *Assume  $m < n$ ,  $q > 1$ . Then*

$$\mathbf{Adv}_{n,m}(q) = \Theta \left( \min \left\{ \frac{q^2}{2^n}, \frac{q}{2^{\frac{n+m}{2}}}, 1 \right\} \right).$$

In particular, note that this implies that the bound (3) is, in general, not tight.

## 2 Notation and Preliminaries

We fix  $0 \leq m < n$  and  $q \geq 1$ . Let  $\mathcal{D}$  be the set of sequences  $(d_\alpha)_{\alpha \in \{0,1\}^{n-m}}$  of non-negative integers such that  $\sum_{\alpha \in \{0,1\}^{n-m}} d_\alpha = q$  and  $d_\alpha \leq 2^m$  for every  $\alpha \in \{0,1\}^{n-m}$ , and let:

$$\Omega := (\{0,1\}^{n-m})^q.$$

We view  $\Omega$  as the set of all possible sequences of replies that the adversary gets for his  $q$  queries. We remark here that in our problem, we may assume that all the queries are fixed and distinct (and hence  $q \leq 2^n$ ). For every  $\omega \in \Omega$  let

$$X(\omega) := \#\{1 \leq i < j \leq q \mid \omega_i = \omega_j\} - \binom{q}{2} \frac{1}{2^{n-m}}.$$

For every  $\omega \in \Omega$ ,  $\alpha \in \{0,1\}^{n-m}$  let

$$d_\alpha(\omega) := \#\{1 \leq i \leq q \mid \omega_i = \alpha\}.$$

For every positive  $\alpha$ , let  $W(0, \alpha) := 1$  and for every positive integer  $k$ ,

$$W(k, \alpha) := \prod_{j=0}^{k-1} \left(1 - \frac{j}{\alpha}\right).$$

For  $\omega \in \Omega$ , let

$$R(\omega) := \frac{\prod_{\alpha \in \{0,1\}^{n-m}} W(d_\alpha(\omega), 2^m)}{W(q, 2^n)}.$$

As in Section 1, consider an adversary that has only query access to a function  $\varphi : \{0,1\}^n \rightarrow \{0,1\}^{n-m}$ , which is either selected uniformly at random from  $\mathcal{F}_{n,n-m}$ , or is  $\text{TRUNC}_{n,m} \circ \pi$ , where  $\pi$  is a permutation of  $\{0,1\}^n$ , chosen uniformly at random. For every  $\omega \in \Omega$ , it easy to verify the following: the probability that  $\omega$  is the actual sequence of replies that the adversary gets for his queries is  $1/|\Omega|$  in the former case, and  $R(\omega)/|\Omega|$  in the latter. Suppose that the adversary uses an algorithm  $\mathcal{A}$ . Let  $S \subseteq \Omega$  be the set of sequences of replies for which  $\mathcal{A}$  outputs 1. Then

$$P_U^{\mathcal{A}}(1) = \sum_{\omega \in S} \frac{1}{|\Omega|}, \quad P_{\text{TRUNC}_{n,m} \circ \pi}^{\mathcal{A}}(1) = \sum_{\omega \in S} \frac{R(\omega)}{|\Omega|},$$

and the advantage of  $\mathcal{A}$  against the PRF  $\text{TRUNC}_{n,m} \circ \pi$  is therefore

$$\left| P_{\text{TRUNC}_{n,m} \circ \pi}^{\mathcal{A}}(1) - P_U^{\mathcal{A}}(1) \right| = \left| \sum_{\omega \in S} \frac{R(\omega) - 1}{|\Omega|} \right|.$$

Assuming the adversary has no computational limitations, we may conclude that

$$\mathbf{Adv}_{n,m}(q) = \mathbf{Adv}_{\text{TRUNC}_{n,m} \circ \pi}(q) = \max_{S \subseteq \Omega} \left| \sum_{\omega \in S} \frac{R(\omega) - 1}{|\Omega|} \right|.$$

Since

$$\left| \sum_{\omega \in S} \frac{R(\omega) - 1}{|\Omega|} \right| = \left| \sum_{\substack{\omega \in S \\ R(\omega) > 1}} \frac{R(\omega) - 1}{|\Omega|} - \sum_{\substack{\omega \in S \\ R(\omega) < 1}} \frac{1 - R(\omega)}{|\Omega|} \right| \leq \max \left\{ \sum_{\substack{\omega \in S \\ R(\omega) > 1}} \frac{R(\omega) - 1}{|\Omega|}, \sum_{\substack{\omega \in S \\ R(\omega) < 1}} \frac{1 - R(\omega)}{|\Omega|} \right\},$$

with equality if  $S = \{\omega \in \Omega \mid R(\omega) > 1\}$  or  $S = \{\omega \in \Omega \mid R(\omega) < 1\}$ , we conclude that

$$\mathbf{Adv}_{n,m}(q) = \sum_{\substack{\omega \in \Omega \\ R(\omega) > 1}} \frac{R(\omega) - 1}{|\Omega|} = \mathbb{E} \max\{R - 1, 0\}, \quad (10)$$

and

$$\mathbf{Adv}_{n,m}(q) = \sum_{\substack{\omega \in \Omega \\ R(\omega) < 1}} \frac{1 - R(\omega)}{|\Omega|} = \mathbb{E} \max\{1 - R, 0\}, \quad (11)$$

where all expectations, here and below, are with respect to the uniform distribution on  $\Omega$ .

### 3 Some useful lemmas

**Lemma 1.** *For every positive  $\alpha$  and positive integer  $k \leq \alpha$ ,*

$$\ln W(k, \alpha) \leq -\frac{k(k-1)}{2\alpha}.$$

*Proof.* Since  $\ln(1-x) \leq -x$  for every  $0 \leq x < 1$ ,

$$\ln W(k, \alpha) = \sum_{j=0}^{k-1} \ln \left( 1 - \frac{j}{\alpha} \right) \leq -\sum_{j=0}^{k-1} \frac{j}{\alpha} = -\frac{k(k-1)}{2\alpha}. \quad \square$$

**Lemma 2.** *For every  $0 \leq x \leq 1/2$ ,*

$$\ln(1-x) \geq -x - x^2.$$

*Proof.* Let  $\varphi(x) := x + 2\ln(4/e)x^2 + \ln(1-x)$ , then

$$\varphi'(x) = 1 + 4\ln(4/e)x - \frac{1}{1-x} = x \left( 4\ln(4/e) - \frac{1}{1-x} \right),$$

and therefore  $\varphi$  increases from 0 to  $1 - \frac{1}{4\ln(4/e)}$  and then decreases. Since  $\varphi(0) = \varphi(1/2) = 0$ , it follows that for every  $0 \leq x \leq 1/2$ ,

$$\ln(1-x) = -x - 2\ln(4/e)x^2 + \varphi(x) \geq -x - 2\ln(4/e)x^2 \geq -x - x^2. \quad \square$$

**Lemma 3.** For every positive  $\alpha$  and positive integer  $k \leq \alpha/2$ ,

$$\ln W(k, \alpha) \geq -\frac{k(k-1)}{2\alpha} - \frac{k^3}{3\alpha^2}.$$

*Proof.* By Lemma 2,

$$\begin{aligned} \ln W(k, \alpha) &= \sum_{j=0}^{k-1} \ln \left( 1 - \frac{j}{\alpha} \right) \geq -\sum_{j=0}^{k-1} \frac{j}{\alpha} - \sum_{j=0}^{k-1} \frac{j^2}{\alpha^2} = \\ &= -\frac{k(k-1)}{2\alpha} - \frac{k(k-1)(2k-1)}{6\alpha^2} \geq -\frac{k(k-1)}{2\alpha} - \frac{k^3}{3\alpha^2}. \quad \square \end{aligned}$$

**Lemma 4.** For every positive  $\alpha$  and positive integer  $k \leq \alpha/2$ ,

$$\ln W(2k, 2\alpha) + \binom{2k}{2} \frac{1}{2\alpha} \geq 2 \left( \ln W(k, \alpha) + \binom{k}{2} \frac{1}{\alpha} \right) - \frac{1}{2} \left( \frac{k}{\alpha} \right)^2$$

*Proof.*

$$\begin{aligned} \frac{W(2k, 2\alpha)^2}{W(k, \alpha)^4} &= \frac{\prod_{j=0}^{2k-1} \left( 1 - \frac{j}{2\alpha} \right)^2}{\prod_{j=0}^{k-1} \left( 1 - \frac{j}{\alpha} \right)^4} = \frac{\prod_{j=0}^{k-1} \left( 1 - \frac{2j}{2\alpha} \right)^2 \left( 1 - \frac{2j+1}{2\alpha} \right)^2}{\prod_{j=0}^{k-1} \left( 1 - \frac{2j}{2\alpha} \right)^4} = \\ &= \frac{\prod_{j=0}^{k-1} \left( 1 - \frac{2j+1}{2\alpha} \right)^2}{\prod_{j=0}^{k-1} \left( 1 - \frac{2j}{2\alpha} \right)^2} = \frac{\prod_{j=0}^{k-1} \left( 1 - \frac{2j+1}{2\alpha} \right)^2}{\prod_{j=0}^{k-1} \left( 1 - \frac{2j}{2\alpha} \right) \left( 1 - \frac{2j+2}{2\alpha} \right)} \left( 1 - \frac{k}{\alpha} \right) \geq 1 - \frac{k}{\alpha}. \end{aligned}$$

Therefore, using Lemma 2,

$$\begin{aligned} \ln W(2k, 2\alpha) + \binom{2k}{2} \frac{1}{2\alpha} - 2 \left( \ln W(k, \alpha) + \binom{k}{2} \frac{1}{\alpha} \right) &= \\ &= \frac{1}{2} \left( \ln \frac{W(2k, 2\alpha)^2}{W(k, \alpha)^4} + \frac{k}{\alpha} \right) \geq \frac{1}{2} \left( \ln \left( 1 - \frac{k}{\alpha} \right) + \frac{k}{\alpha} \right) \geq -\frac{1}{2} \left( \frac{k}{\alpha} \right)^2. \quad \square \end{aligned}$$

**Lemma 5.** Suppose  $q$  is a power of 2. Then for every  $(d_\alpha)_{\alpha \in \{0,1\}^{n-m}} \in \mathcal{D}$ ,

$$\sum_{\alpha \in \{0,1\}^{n-m}} \ln W(d_\alpha, 2^m) + \binom{d_\alpha}{2} \frac{1}{2^m} \leq \begin{cases} 0 & q < 2^{n-m} \\ 2^{n-m} \left( \ln W \left( \frac{q}{2^{n-m}}, 2^m \right) + \left( \frac{q}{2^{n-m}} \right) \frac{1}{2^m} \right) & q \geq 2^{n-m} \end{cases}$$

*Proof.* For every  $(d_\alpha)_{\alpha \in \{0,1\}^{n-m}} \in \mathcal{D}$ , let

$$F((d_\alpha)_{\alpha \in \{0,1\}^{n-m}}) := \sum_{\alpha \in \{0,1\}^{n-m}} \ln W(d_\alpha, 2^m) + \binom{d_\alpha}{2} \frac{1}{2^m}.$$

Suppose  $d_{\alpha_1} + 1 \leq d_{\alpha_2} - 1$  for some  $\alpha_1, \alpha_2 \in \{0,1\}^{n-m}$ . Define  $(\tilde{d}_\alpha)_{\alpha \in \{0,1\}^{n-m}}$  by:

$$\tilde{d}_\alpha := \begin{cases} d_\alpha & \alpha \notin \{\alpha_1, \alpha_2\} \\ d_{\alpha_1} + 1 & \alpha = \alpha_1 \\ d_{\alpha_2} - 1 & \alpha = \alpha_2 \end{cases}$$

Then  $(\tilde{d}_\alpha)_{\alpha \in \{0,1\}^{n-m}} \in \mathcal{D}$  and, since the function  $x \mapsto \ln(1-x) + x$  is strictly decreasing in  $[0,1)$ ,

$$\begin{aligned} F((\tilde{d}_\alpha)_{\alpha \in \{0,1\}^{n-m}}) - F((d_\alpha)_{\alpha \in \{0,1\}^{n-m}}) &= \\ &= \left( \ln W(\tilde{d}_{\alpha_1}, 2^m) - \ln W(d_{\alpha_1}, 2^m) \right) + \left( \binom{\tilde{d}_{\alpha_1}}{2} - \binom{d_{\alpha_1}}{2} \right) \frac{1}{2^m} + \\ &\quad + \left( \ln W(\tilde{d}_{\alpha_2}, 2^m) - \ln W(d_{\alpha_2}, 2^m) \right) + \left( \binom{\tilde{d}_{\alpha_2}}{2} - \binom{d_{\alpha_2}}{2} \right) \frac{1}{2^m} = \\ &= \left( \ln W(d_{\alpha_1} + 1, 2^m) - \ln W(d_{\alpha_1}, 2^m) \right) + \left( \binom{d_{\alpha_1} + 1}{2} - \binom{d_{\alpha_1}}{2} \right) \frac{1}{2^m} + \\ &\quad + \left( \ln W(d_{\alpha_2} - 1, 2^m) - \ln W(d_{\alpha_2}, 2^m) \right) + \left( \binom{d_{\alpha_2} - 1}{2} - \binom{d_{\alpha_2}}{2} \right) \frac{1}{2^m} = \\ &= \left( \ln \left( 1 - \frac{d_{\alpha_1}}{2^m} \right) + \frac{d_{\alpha_1}}{2^m} \right) - \left( \ln \left( 1 - \frac{d_{\alpha_2}}{2^m} \right) + \frac{d_{\alpha_2}}{2^m} \right) > 0 \end{aligned}$$

Therefore, if  $q \geq 2^{n-m}$  then the maximum of  $F$  in  $\mathcal{D}$  is attained at the sequence  $(d_\alpha)_{\alpha \in \{0,1\}^{n-m}}$  such that  $d_\alpha = q/2^{n-m}$  for every  $\alpha \in \{0,1\}^{n-m}$ . If  $q < 2^{n-m}$  then the maximum of  $F$  in  $\mathcal{D}$  is attained at any  $(d_\alpha)_{\alpha \in \{0,1\}^{n-m}} \in \mathcal{D}$  for which  $d_\alpha \leq 1$  for every  $\alpha \in \{0,1\}^{n-m}$ . Therefore, for every  $(d_\alpha)_{\alpha \in \{0,1\}^{n-m}} \in \mathcal{D}$ ,

$$\begin{aligned} \sum_{\alpha \in \{0,1\}^{n-m}} \ln W(d_\alpha, 2^m) + \binom{d_\alpha}{2} \frac{1}{2^m} &= F((d_\alpha)_{\alpha \in \{0,1\}^{n-m}}) \leq \\ &\leq \begin{cases} 0 & q < 2^{n-m} \\ 2^{n-m} \left( \ln W\left(\frac{q}{2^{n-m}}, 2^m\right) + \binom{\frac{q}{2^{n-m}}}{2} \frac{1}{2^m} \right) & q \geq 2^{n-m} \end{cases} \quad \square \end{aligned}$$

**Lemma 6.** Suppose  $q$  is a power of 2. Then

$$R \leq e^{\frac{1}{2} \cdot \frac{q^2}{2^{n+m}} - \frac{1}{2^m} X}.$$



*Proof.* Let  $\omega \in \Omega$ . If  $d_\alpha(\omega) > 2^m$  for some  $\alpha \in \{0, 1\}^{n-m}$ , then surely

$$R(\omega) = 0 < e^{\frac{1}{2} \cdot \frac{q^2}{2^{n+m}} - \frac{1}{2^m} X(\omega)}.$$

We therefore assume that  $d_\alpha(\omega) \leq 2^m$  for every  $\alpha \in \{0, 1\}^{n-m}$ , and hence  $(d_\alpha(\omega))_{\alpha \in \{0, 1\}^{n-m}} \in \mathcal{D}$ . Note that

$$X(\omega) + \binom{q}{2} \frac{1}{2^{n-m}} = \sum_{\alpha \in \{0, 1\}^{n-m}} \binom{d_\alpha(\omega)}{2},$$

hence

$$\ln R(\omega) + \ln W(q, 2^n) + \frac{1}{2^m} X(\omega) + \binom{q}{2} \frac{1}{2^n} = \sum_{\alpha \in \{0, 1\}^{n-m}} \ln W(d_\alpha(\omega), 2^m) + \binom{d_\alpha(\omega)}{2} \frac{1}{2^m}.$$

If  $q < 2^{n-m}$  then by Lemma 5 and Lemma 3 we conclude that

$$\begin{aligned} \ln R(\omega) &\leq -\ln W(q, 2^n) - \frac{1}{2^m} X(\omega) - \binom{q}{2} \frac{1}{2^n} = -\left(\ln W(q, 2^n) + \frac{q(q-1)}{2 \cdot 2^n}\right) - \frac{1}{2^m} X(\omega) \leq \\ &\leq \frac{q^3}{3 \cdot 2^{2n}} - \frac{1}{2^m} X(\omega) = \frac{1}{6} \cdot \frac{2q}{2^{n-m}} \cdot \frac{q^2}{2^{n+m}} - \frac{1}{2^m} X(\omega) < \frac{1}{2} \cdot \frac{q^2}{2^{n+m}} - \frac{1}{2^m} X(\omega). \end{aligned}$$

If  $q \geq 2^{n-m}$  then by Lemma 5 and a repetitive use of Lemma 4,

$$\begin{aligned} \ln R(\omega) + \ln W(q, 2^n) + \frac{1}{2^m} X(\omega) + \binom{q}{2} \frac{1}{2^n} &\leq 2^{n-m} \left( \ln W\left(\frac{q}{2^{n-m}}, 2^m\right) + \binom{\frac{q}{2^{n-m}}}{2} \frac{1}{2^m} \right) \leq \\ &\leq \ln W(q, 2^n) + \binom{q}{2} \frac{1}{2^n} + (2^{n-m} - 1) \frac{1}{2} \left(\frac{q}{2^n}\right)^2, \end{aligned}$$

hence

$$\ln R(\omega) \leq (2^{n-m} - 1) \frac{1}{2} \left(\frac{q}{2^n}\right)^2 - \frac{1}{2^m} X(\omega) \leq \frac{1}{2} \cdot \frac{q^2}{2^{n+m}} - \frac{1}{2^m} X(\omega). \quad \square$$

**Lemma 7.**

$$EX = 0,$$

$$EX^2 = \binom{q}{2} \frac{1}{2^{n-m}} \left(1 - \frac{1}{2^{n-m}}\right),$$

$$EX^3 = 6 \binom{q}{3} \frac{1}{2^{2(n-m)}} \left(1 - \frac{1}{2^{n-m}}\right) + \binom{q}{2} \frac{1}{2^{n-m}} \left(1 - \frac{1}{2^{n-m}}\right) \left(1 - \frac{2}{2^{n-m}}\right),$$

$$\begin{aligned} EX^4 = & 18 \binom{q}{4} \frac{1}{2^{2(n-m)}} \left(1 - \frac{1}{2^{n-m}}\right) \left(1 + \frac{3}{2^{n-m}}\right) + 54 \binom{q}{3} \frac{1}{2^{2(n-m)}} \left(1 - \frac{1}{2^{n-m}}\right) \left(1 - \frac{5}{3 \cdot 2^{n-m}}\right) + \\ & + \binom{q}{2} \frac{1}{2^{n-m}} \left(1 - \frac{1}{2^{n-m}}\right) \left(1 - \frac{3}{2^{n-m}} + \frac{3}{2^{2(n-m)}}\right). \end{aligned}$$

*Proof.* Denote  $p := 1/2^{n-m}$  and  $\mathcal{E} := \{\{i, j\} \mid 1 \leq i < j \leq q\}$ . Note that

$$X = \sum_{e \in \mathcal{E}} (X_e - p),$$

where for every  $1 \leq i < j \leq q$ ,  $X_{\{i,j\}}$  is the indicator function of the event  $\{\omega_i = \omega_j\}$ , whose probability is clearly  $p$ . Therefore, for every  $e \in \mathcal{E}$ ,

$$\mathbb{E}X_e = p, \quad (12)$$

hence, by linearity of expectation,

$$\mathbb{E}X = \sum_{e \in \mathcal{E}} (\mathbb{E}X_e - p) = 0.$$

For every  $e_1, e_2 \in \mathcal{E}$ , clearly

$$\mathbb{E}X_{e_1}X_{e_2} = \begin{cases} p & e_1 = e_2 \\ p^2 & e_1 \neq e_2, \end{cases} \quad (13)$$

hence, using (12),

$$\mathbb{E}(X_{e_1} - p)(X_{e_2} - p) = \mathbb{E}X_{e_1}X_{e_2} - p(\mathbb{E}X_{e_1} + \mathbb{E}X_{e_2}) + p^2 = \mathbb{E}X_{e_1}X_{e_2} - p^2 = \begin{cases} p(1-p) & e_1 = e_2 \\ 0 & e_1 \neq e_2, \end{cases}$$

and, again by linearity of expectation, we get

$$\mathbb{E}X^2 = \mathbb{E} \sum_{e_1, e_2 \in \mathcal{E}} (X_{e_1} - p)(X_{e_2} - p) = \sum_{e_1, e_2 \in \mathcal{E}} \mathbb{E}(X_{e_1} - p)(X_{e_2} - p) = \sum_{e \in \mathcal{E}} p(1-p) = \binom{q}{2} p(1-p).$$

For every  $e_1, e_2, e_3 \in \mathcal{E}$ ,

$$\mathbb{E}X_{e_1}X_{e_2}X_{e_3} = \begin{cases} p & e_1 = e_2 = e_3 \\ p^2 & |\{e_1, e_2, e_3\}| = 2 \text{ or } |e_1 \cup e_2 \cup e_3| = 3 \\ p^3 & \text{otherwise,} \end{cases} \quad (14)$$

hence, using (12) and (13), we conclude that

$$\begin{aligned} \mathbb{E}(X_{e_1} - p)(X_{e_2} - p)(X_{e_3} - p) &= \\ &= \mathbb{E}X_{e_1}X_{e_2}X_{e_3} - p(\mathbb{E}X_{e_1}X_{e_2} + \mathbb{E}X_{e_1}X_{e_3} + \mathbb{E}X_{e_2}X_{e_3}) + p^2(\mathbb{E}X_{e_1} + \mathbb{E}X_{e_2} + \mathbb{E}X_{e_3}) - p^3 = \\ &= \mathbb{E}X_{e_1}X_{e_2}X_{e_3} - p(\mathbb{E}X_{e_1}X_{e_2} + \mathbb{E}X_{e_1}X_{e_3} + \mathbb{E}X_{e_2}X_{e_3}) + 2p^3 = \\ &= \begin{cases} p(1-p)(1-2p) & e_1 = e_2 = e_3 \\ p^2(1-p) & \text{the graph with edges } e_1, e_2, e_3 \text{ forms a cycle (triangle)} \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

and, as before, we get

$$\mathbb{E}X^3 = 6 \binom{q}{3} p^2(1-p) + \binom{q}{2} p(1-p)(1-2p).$$

Finally, for every  $e_1, e_2, e_3, e_4 \in \mathcal{E}$ ,

$$\mathbb{E}X_{e_1}X_{e_2}X_{e_3}X_{e_4} = \begin{cases} p & e_1 = e_2 = e_3 = e_4 \\ p^2 & |\{e_1, e_2, e_3, e_4\}| = 2 \text{ or } |e_1 \cup e_2 \cup e_3 \cup e_4| = 3 \\ p^4 & \text{the graph with edges } e_1, e_2, e_3, e_4 \text{ forms a forest} \\ p^3 & \text{otherwise,} \end{cases}$$

hence, using (12), (13) and (14), we can conclude that

$$\begin{aligned} & \mathbb{E}(X_{e_1} - p)(X_{e_2} - p)(X_{e_3} - p)(X_{e_4} - p) = \\ & = \mathbb{E}X_{e_1}X_{e_2}X_{e_3}X_{e_4} - p(\mathbb{E}X_{e_1}X_{e_2}X_{e_3} + \mathbb{E}X_{e_1}X_{e_2}X_{e_4} + \mathbb{E}X_{e_1}X_{e_3}X_{e_4} + \mathbb{E}X_{e_2}X_{e_3}X_{e_4}) + \\ & \quad + p^2(\mathbb{E}X_{e_1}X_{e_2} + \mathbb{E}X_{e_1}X_{e_3} + \mathbb{E}X_{e_1}X_{e_4} + \mathbb{E}X_{e_2}X_{e_3} + \mathbb{E}X_{e_2}X_{e_4} + \mathbb{E}X_{e_3}X_{e_4}) - \\ & \quad - p^3(\mathbb{E}X_{e_1} + \mathbb{E}X_{e_2} + \mathbb{E}X_{e_3} + \mathbb{E}X_{e_4}) + p^4 = \\ & = \mathbb{E}X_{e_1}X_{e_2}X_{e_3}X_{e_4} - p(\mathbb{E}X_{e_1}X_{e_2}X_{e_3} + \mathbb{E}X_{e_1}X_{e_2}X_{e_4} + \mathbb{E}X_{e_1}X_{e_3}X_{e_4} + \mathbb{E}X_{e_2}X_{e_3}X_{e_4}) + \\ & \quad + p^2(\mathbb{E}X_{e_1}X_{e_2} + \mathbb{E}X_{e_1}X_{e_3} + \mathbb{E}X_{e_1}X_{e_4} + \mathbb{E}X_{e_2}X_{e_3} + \mathbb{E}X_{e_2}X_{e_4} + \mathbb{E}X_{e_3}X_{e_4}) - 3p^4 = \\ & = \begin{cases} p(1-p)(1-3p+3p^2) & e_1 = e_2 = e_3 = e_4 \\ p^2(1-p)(1-2p) & |\{e_1, e_2, e_3, e_4\}| = 3 \text{ and } |e_1 \cup e_2 \cup e_3 \cup e_4| = 3 \text{ }^{(*)} \\ p^2(1-p)^2 & \forall 1 \leq i \leq 4 : |\{1 \leq j \leq 4 \mid e_i = e_j\}| = 2 \text{ }^{(**)} \\ p^3(1-p) & \text{the graph with edges } e_1, e_2, e_3, e_4 \text{ forms a cycle (quadrilateral)} \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

and therefore

$$\begin{aligned} \mathbb{E}X^4 &= 72 \binom{q}{4} p^3(1-p) + \left(18 \binom{q}{4} + 18 \binom{q}{3}\right) p^2(1-p)^2 + \\ & \quad + 36 \binom{q}{3} p^2(1-p)(1-2p) + \binom{q}{2} p(1-p)(1-3p+3p^2) = \\ & = 18 \binom{q}{4} p^2(1-p)(1+3p) + 54 \binom{q}{3} p^2(1-p) \left(1 - \frac{5}{3}p\right) + \binom{q}{2} p(1-p)(1-3p+3p^2). \quad \square \end{aligned}$$

**Lemma 8.** If  $q > 2^{\frac{n-m}{2}+8}$  then

$$\mathbb{E}X^4 < \frac{q^2(q-1)^2}{2^{2(n-m)}}.$$

*Proof.* First note that

$$\left(1 - \frac{1}{2^{n-m}}\right) \left(1 + \frac{3}{2^{n-m}}\right) \leq \left(1 - \frac{1}{4}\right) \left(1 + \frac{3}{4}\right) = \frac{21}{16}$$

---

<sup>(\*)</sup> i.e., the graph with edges  $e_1, e_2, e_3, e_4$  forms a triangle with a double edge.

<sup>(\*\*)</sup> in words: the four edges are divided to two pairs of double edges.

(this follows by direct computation if  $n-m=1$ , and for  $n-m \geq 2$  since the function  $x \mapsto (1-x)(1+3x)$  is increasing on  $[0, 1/3]$ ), therefore

$$18 \binom{q}{4} \frac{1}{2^{2(n-m)}} \left(1 - \frac{1}{2^{n-m}}\right) \left(1 + \frac{3}{2^{n-m}}\right) < 18 \cdot \frac{q^2(q-1)^2}{24} \cdot \frac{1}{2^{2(n-m)}} \cdot \frac{21}{16} = \left(1 - \frac{1}{2^6}\right) \frac{q^2(q-1)^2}{2^{2(n-m)}}.$$

Next, note that

$$\left(1 - \frac{1}{2^{n-m}}\right) \left(1 - \frac{5}{3 \cdot 2^{n-m}}\right) < \frac{1}{4} 2^{\frac{n-m}{2}}$$

(it is clear if  $n-m \geq 4$ , and easy to verify if  $1 \leq n-m \leq 3$ ). Therefore,

$$\begin{aligned} 54 \binom{q}{3} \frac{1}{2^{2(n-m)}} \left(1 - \frac{1}{2^{n-m}}\right) \left(1 - \frac{5}{3 \cdot 2^{n-m}}\right) &< 54 \binom{q}{3} \frac{1}{2^{2(n-m)}} \cdot \frac{1}{4} 2^{\frac{n-m}{2}} < \\ &< 9 \cdot \frac{q(q-1)^2}{2^{2(n-m)}} \cdot \frac{1}{4} 2^{\frac{n-m}{2}} < 9 \cdot \frac{q(q-1)^2}{2^{2(n-m)}} \cdot \frac{1}{4} \cdot \frac{q}{2^8} = \frac{9}{2^{10}} \cdot \frac{q^2(q-1)^2}{2^{2(n-m)}}. \end{aligned}$$

In addition

$$\binom{q}{2} \frac{1}{2^{n-m}} \left(1 - \frac{1}{2^{n-m}}\right) \left(1 - \frac{3}{2^{n-m}} + \frac{3}{2^{2(n-m)}}\right) < \frac{1}{2} \cdot \frac{q(q-1)}{2^{n-m}} < \frac{1}{2^{17}} \cdot \frac{q^2(q-1)^2}{2^{2(n-m)}}.$$

Therefore, by Lemma 7 we get that

$$\mathbb{E}X^4 < \left(1 - \frac{1}{2^6} + \frac{9}{2^{10}} + \frac{1}{2^{17}}\right) \frac{q^2(q-1)^2}{2^{2(n-m)}} < \frac{q^2(q-1)^2}{2^{2(n-m)}}. \quad \square$$

**Lemma 9.** Suppose  $Y$  is a random variable bounded above by a positive real number  $M$ , then

$$\Pr(Y > 0) \geq \frac{\mathbb{E}Y}{M}.$$

*Proof.* The random variable  $M - Y$  is non-negative, hence by Markov's inequality

$$\Pr(Y \leq 0) = \Pr(M - Y \geq M) \leq \frac{\mathbb{E}(M - Y)}{M} = \frac{M - \mathbb{E}Y}{M} = 1 - \frac{\mathbb{E}Y}{M}$$

and therefore

$$\Pr(Y > 0) = 1 - \Pr(Y \leq 0) \geq \frac{\mathbb{E}Y}{M}. \quad \square$$

**Lemma 10.** If  $q > 2^{\frac{n-m}{2}+8}$  then

$$\Pr\left(X > \frac{1}{10} \cdot \frac{\sqrt{q(q-1)}}{2^{(n-m)/2}}\right) > \frac{1}{400}$$

*Proof.* For every real  $x$ , let

$$\varphi(x) := -\left(x + \frac{5}{2}\right)^2 \left(x - \frac{1}{10}\right) (x - 5) = -x^4 + \frac{1}{10}x^3 + \frac{75}{4}x^2 + \frac{235}{8}x - \frac{25}{8}.$$

Since for every real  $x$ ,

$$\varphi'(x) = -4 \left( x + \frac{5}{2} \right) \left( x - \frac{103 - \sqrt{29409}}{80} \right) \left( x - \frac{103 + \sqrt{29409}}{80} \right),$$

we get that for every real  $x$ ,

$$\varphi(x) \leq \varphi \left( \frac{103 + \sqrt{29409}}{80} \right) < 200.$$

By Lemma 7 and Lemma 8:

$$\begin{aligned} \mathbb{E} \varphi \left( \frac{2^{(n-m)/2}}{\sqrt{q(q-1)}} X \right) &= - \frac{2^{2(n-m)}}{q^2(q-1)^2} \mathbb{E} X^4 + \frac{1}{10} \cdot \frac{2^{3(n-m)/2}}{q^{3/2}(q-1)^{3/2}} \mathbb{E} X^3 + \\ &\quad + \frac{75}{4} \cdot \frac{2^{n-m}}{q(q-1)} \mathbb{E} X^2 + \frac{235}{8} \cdot \frac{2^{(n-m)/2}}{\sqrt{q(q-1)}} \mathbb{E} X - \frac{25}{8} = \\ &= - \frac{2^{2(n-m)}}{q^2(q-1)^2} \mathbb{E} X^4 + \frac{1}{10} \cdot \frac{2^{3(n-m)/2}}{q^{3/2}(q-1)^{3/2}} \mathbb{E} X^3 + \frac{75}{4} \cdot \frac{1}{2} \left( 1 - \frac{1}{2^{n-m}} \right) - \frac{25}{8} \geq \\ &\geq -1 + \frac{75}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} - \frac{25}{8} > \frac{1}{2}. \end{aligned}$$

Hence, by Lemma 9 we get:

$$\begin{aligned} \Pr \left( X > \frac{1}{10} \cdot \frac{\sqrt{q(q-1)}}{2^{(n-m)/2}} \right) &\geq \Pr \left( \varphi \left( \frac{2^{(n-m)/2}}{\sqrt{q(q-1)}} X \right) > 0 \right) \geq \\ &\geq \frac{1}{200} \mathbb{E} \varphi \left( \frac{2^{(n-m)/2}}{\sqrt{q(q-1)}} X \right) > \frac{1}{200} \cdot \frac{1}{2} = \frac{1}{400}. \quad \square \end{aligned}$$

## 4 Proof of Theorem 1

In this section we prove our main result. We first address the regime  $1 < q \leq 2^{\frac{n-m}{2}+8}$ , in which

$$\min \left\{ \frac{q^2}{2^n}, \frac{q}{2^{\frac{n+m}{2}}}, 1 \right\} = \Theta \left( \frac{q^2}{2^n} \right).$$

**Proposition 1.** *If  $1 < q \leq 2^{\frac{n-m}{2}+8}$  then*

$$\mathbf{Adv}_{n,m}(q) = \Omega \left( \frac{q^2}{2^n} \right).$$

*Proof.* Assume first, in addition, that  $q \leq 2^{n-m-1}$ . Let

$$S := \{ \omega \in \Omega \mid \forall \alpha \in \{0, 1\}^{n-m} : d_\alpha(\omega) \leq 1 \}.$$

By Lemma 1, for every  $\omega \in S$

$$R(\omega) - 1 = \frac{1}{W(q, 2^n)} - 1 \geq e^{\frac{q(q-1)/2}{2^n}} - 1 \geq \frac{q(q-1)/2}{2^n},$$

and by Lemma 3,

$$\Pr(S) = W(q, 2^{n-m}) = \Omega(1).$$

Hence, by (10),

$$\mathbf{Adv}_{n,m}(q) = \mathbb{E} \max\{R - 1, 0\} \geq \Pr(S) \frac{q(q-1)/2}{2^n} = \Omega\left(\frac{q^2}{2^n}\right).$$

Now, if  $2^{n-m-1} < q \leq 2^{\frac{n-m}{2}+8}$ , then by what we already proved

$$\mathbf{Adv}_{n,m}(q) \geq \mathbf{Adv}_{n,m}(2^{n-m-1}) = \Omega\left(\frac{(2^{n-m-1})^2}{2^n}\right) = \Omega\left(\frac{q^2}{2^n}\right). \quad \square$$

We now address the regime  $2^{\frac{n-m}{2}+8} < q \leq 2^{\frac{n+m}{2}-3}$ , in which

$$\min\left\{\frac{q^2}{2^n}, \frac{q}{2^{\frac{n+m}{2}}}, 1\right\} = \Theta\left(\frac{q}{2^{\frac{n+m}{2}}}\right).$$

**Proposition 2.** Assume that  $2^{\frac{n-m}{2}+8} < q \leq 2^{\frac{n+m}{2}-3}$ . Then

$$\mathbf{Adv}_{n,m}(q) = \Omega\left(\frac{q}{2^{\frac{n+m}{2}}}\right).$$

*Proof.* With no loss of generality we may assume  $q$  is a power of 2. If  $X(\omega) > \frac{1}{10} \cdot \frac{\sqrt{q(q-1)}}{2^{(n-m)/2}}$  then

$$\begin{aligned} \frac{1}{2} \cdot \frac{q^2}{2^{n+m}} - \frac{1}{2^m} X(\omega) &< \frac{1}{2} \cdot \frac{q^2}{2^{n+m}} - \frac{1}{2^m} \cdot \frac{1}{10} \cdot \frac{\sqrt{q(q-1)}}{2^{\frac{n-m}{2}}} = \\ &= -\frac{1}{10} \left(1 - \frac{5}{\sqrt{1 - \frac{1}{q}}} \cdot \frac{q}{2^{\frac{n+m}{2}}}\right) \frac{\sqrt{q(q-1)}}{2^{\frac{n+m}{2}}} < -\frac{1}{27} \cdot \frac{\sqrt{q(q-1)}}{2^{\frac{n+m}{2}}}, \end{aligned}$$

hence, by Lemma 6,

$$1 - R(\omega) > 1 - e^{-\frac{1}{27} \cdot \frac{\sqrt{q(q-1)}}{2^{(n+m)/2}}}.$$

Therefore, by (11) and Lemma 10,

$$\mathbf{Adv}_{n,m}(q) = \mathbb{E} \max\{1 - R, 0\} > \frac{1}{400} \left(1 - e^{-\frac{1}{27} \cdot \frac{\sqrt{q(q-1)}}{2^{(n+m)/2}}}\right) = \Omega\left(\frac{q}{2^{\frac{n+m}{2}}}\right). \quad \square$$

Now we can prove Theorem 1.

*Proof of Theorem 1.* The upper bound was already demonstrated in the introduction, so we only need to show that

$$\mathbf{Adv}_{n,m}(q) = \Omega\left(\min\left\{\frac{q^2}{2^n}, \frac{q}{2^{\frac{n+m}{2}}}, 1\right\}\right).$$

If  $1 < q \leq 2^{\frac{n-m}{2}+8}$  then by Proposition 1

$$\mathbf{Adv}_{n,m}(q) = \Omega\left(\frac{q^2}{2^n}\right).$$

If  $2^{\frac{n-m}{2}+8} < q \leq 2^{\frac{n+m}{2}-3}$  then by Proposition 2

$$\mathbf{Adv}_{n,m}(q) = \Omega\left(\frac{q}{2^{\frac{n+m}{2}}}\right).$$

Finally, if  $q > 2^{\frac{n+m}{2}-3}$  then by Proposition 2

$$\mathbf{Adv}_{n,m}(q) \geq \mathbf{Adv}_{n,m}\left(2^{\frac{n+m}{2}-3}\right) = \Omega\left(\frac{2^{\frac{n+m}{2}-3}}{2^{\frac{n+m}{2}}}\right) = \Omega(1). \quad \square$$

## 5 Conclusions

Theorem 1 settled Problem 1 by showing that the upper bound (9) is tight for every  $q > 1$ .

Our interpretation is that the truncated permutation PRF is a tool that can use a *single* permutation (selected uniformly and random) of  $\{0,1\}^n$  to produce a pseudo random bits stream whose length exceeds the birthday bound.

For a concrete example consider  $n = 128$ ,  $m = 64$ , and  $q = 2^{64}$ . The simplified bound in (8) indicates it is possible to generate stream with length of  $2^{67}$  bytes with an upper bound of  $2^{-32}$  on the distinguishing advantage, where this bound cannot be improved fundamentally. In a real life context, consider the 128-bit block cipher AES. It is commonly believed (per its design goal and the analyses it went through) to be indistinguishable from a random permutation of  $\{0,1\}^{128}$  if it is used with a 128-bit key that is chosen uniformly at random, even if an adversary can see a very large amount of samples. Under this assumption and the above results, it is possible to use 64-bit truncated outputs of AES as a Beyond-Birthday-Bound PRF, with indistinguishability margin  $\sim 2^{32}$ . The stream generation rate of this Beyond-Birthday-Bound PRF is half the throughput of AES, and is very high on modern processors due to the existence of the AES instructions (1.3 cycles per byte on the latest Intel processors, micro-architecture codename Skylake).

## Acknowledgments

We thank Ron Peled for fruitful discussion, and Ben Morris for introducing us to the work of Stam.

This research was supported by the the PQCRYPTO project, which is partially funded by the European Commission Horizon 2020 research Programme, grant #645622, and by the ISRAEL SCIENCE FOUNDATION (grant No. 1018/16).

## References

- [1] M. Bellare, R. Impagliazzo, "A tool for obtaining tighter security analyses of pseudo-random function based constructions, with applications to PRP to PRF conversion", ePrint 1999/024, <http://eprint.iacr.org/1999/024> (1999).
- [2] S. Gilboa, S. Gueron, Distinguishing a truncated random permutation from a random function, unpublished manuscript, available at [arXiv:1508.00462](https://arxiv.org/abs/1508.00462).
- [3] S. Gilboa, S. Gueron and B. Morris, How many queries are needed to distinguish a truncated random permutation from a random function? Preprint available at [arXiv:1412.5204](https://arxiv.org/abs/1412.5204).
- [4] C. Hall, D. Wagner, J. Kelsey, B. Schneier, Building prfs from prps, in: Proceedings of CRYPTO-98: Advances in Cryptography, Springer Verlag, 1998, pp. 370-389.
- [5] A. J. Stam, Distance between sampling with and without replacement, *Statist. Neerlandica* **32** (1978), no. 2, 81–91.